DOCUMENT RESUME

ED 183 877 CE 024 497

AUTHOR McCaslin, N. L.; Walker, Jerry P.
TITLE Career Education Measurement Handbooks. A Guide for Improving Locally Developed Career Education Measures. Research & Development Series No. 167.
INSTITUTION Ohio State Univ., Columbus. National Center for Research in Vocational Education.
SPONS AGENCY National Inst. of Education (DHEW), Washington, D.C.
PUB DATE 79
CONTRACT NE-C-00-3-0079
NOTE 60p.; For related documents see CE 024 495-499.
AVAILABLE FROM National Center Publications, National Center for Research in Vocational Education, The Ohio State University, 1960 Kenny Road, Columbus, OH 43210 ($2.80; $30.00 for complete set)

EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS *Career Education; *Educational Assessment; Guidelines; *Measures (Individuals); Readability; Sex Fairness; *Test Construction; Testing; Test Reliability; Test Selection; Test Validity

ABSTRACT

This document is the third volume in a set of five Career Education Measurement Handbooks intended to help local education personnel in measurement and evaluation. Divided into seven major sections, this handbook is designed to provide practical guidelines for improving career education instruments developed at the local level. It is primarily intended for people who lack training in evaluation and measurement but are responsible for this type of work. A brief introduction providing a checklist for rating career education instruments precedes the first section, which includes guidelines for planning and designing career education instruments. Sections 2 and 3 present information on checking reliability and determining validity, respectively. Eliminating stereotypes is the topic of section 4. Section 5 presents guidelines for devising the format. Section 6 discusses readability, and section 7 provides guidelines for examining content validity and internal consistency reliability. (BM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED183877

Research & Development Series No. 167

CAREER EDUCATION MEASUREMENT HANDBOOKS.

A GUIDE FOR IMPROVING LOCALLY DEVELOPED CAREER EDUCATION MEASURES

N. L. McCaslin
Jerry P. Walker

The National Center for Research in Vocational Education
The Ohio State University
Columbus, Ohio

1979

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

CE 024 497

THE NATIONAL CENTER MISSION STATEMENT

The National Center for Research in Vocational Education's mission is to increase the ability of diverse agencies, institutions, and organizations to solve educational problems relating to individual career planning, preparation, and progression. The National Center fulfills its mission by:

- Generating knowledge through research
- Developing educational programs and products
- Evaluating individual program needs and outcomes
- Installing educational programs and products
- Operating information systems and services
- Conducting leadership development and training programs

The work upon which this publication is based was performed pursuant to Contract NE-C-00-3-0079 with the National Institute of Education, Department of Health, Education, and Welfare. Contractors undertaking such projects under government sponsorship are encouraged to freely express their judgement in professional and technical matters. Points of view or opinions do not, therefore, necessarily represent official National Institute of Education position or policy.

U.S. DEPARTMENT OF
HEALTH, EDUCATION, AND WELFARE

National Institute of Education

FOREWORD

Educators have often been criticized for viewing their role merely as dispensers of knowledge and for showing a lack of concern with the application of this knowledge to their students' future lives. However, the general public has begun to accept the idea that the educational system has a responsibility to assist *all* individuals in making orderly transitions to the world of work. Additionally, during the last decade, a number of innovative activities, projects, and programs have been developed at the federal, state, and local levels in which personnel are attempting to link education and work. Among these programs and projects are the following examples: career education, Experience-Based Career Education (EBCE), Part D exemplary projects and other projects in vocational education, Title IV-C of the Elementary and Secondary Education Act (ESEA), Fund for the Improvement of Post Secondary Education (FIPSE), and Title I and III of the Comprehensive Employment and Training Act (CETA).

Personnel associated with education and work programs are becoming increasingly aware of the need for information related to accountability and needed improvements. Many reports—some informal, some in the literature—indicate a wide and exciting variety of approaches to evaluating programs that link education and work. However, the need exists for practitioners to become acquainted with evaluation ideas and materials available for particular situations.

Recognizing these trends, the Education and Work Group of the National Institute of Education (NIE) contracted with the National Center for Research in Vocational Education (NCRVE) to develop the *Career Education Measurement Series.* The series includes five comprehensive "user oriented" handbooks intended to help local education personnel in measurement and evaluation. The handbooks in the series are:

- *Assessing Experiential Learning in Career Education*
- *Career Education Measures: A Compendium of Evaluation Instruments*
- *Improving the Accountability of Career Education Programs: Evaluation Guidelines and Checklists*
- *A Guide for Improving Locally Developed Career Education Measures*
- *Using Systematic Observation Techniques in Evaluating Career Education*

This handbook, *A Guide for Improving Locally Developed Career Education Measures,* is designed to provide practical guidelines for improving career education instruments developed at the local level. The handbook is primarily designed for people who lack training in evaluation and measurement but are responsible for this type of work.

The Center is particularly indebted to Dr. N. L. McCaslin, Project Director and Dr. Jerry P. Walker, Associate Director for Evaluation at The Center, who prepared this handbook. Additionally, appreciation is extended to Dr. Michael Neuman, Assistant Professor of English at Capital University,

who edited the handbook, and to Dr. Kay Adams, Research Specialist, for reviewing developmental versions of the manuscript. A special note of thanks is extended to Dr. Ronald Bucknam who originally conceived this handbook series and continued his involvement through development as the Project Officer for the National Institute of Education. Valuable advice in the design and scope of the handbook series was received from an advisory committee composed of Dr. Robert Ebel, Michigan State University; Dr. Margaret Fergueron State Director of Career Education in Florida; and Ms. Deede Sharpe, Georgia Department of Education.

In an attempt to make this handbook truly "user oriented," a user trial was conducted prior to publication of the handbook. The National Center is indebted to those career education practitioners who participated in the user trial. Their valuable assistance has greatly enhanced the utility of this handbook. These individuals included:

Roland J. Cross	Orono, Wisconsin
Caroll Curtis	Pennsylvania Department of Education
Pat Doherty	New Jersey Department of Education
H. Stephen Hewlett	Valparaiso, Indiana
Ellen Meister	Madison, Wisconsin
David G. Minnis	Edinboro, Pennsylvania
Bernard Novick	Woodbridge, New Jersey
Harry E. Owen	Winchester, Kentucky
Carl W. Pepperman	Lewisburg, Pennsylvania
Allen I. Stessmann	Sheboygan, Wisconsin
Rose Trachtenberg	Hammond, Indiana

Finally, a special note of appreciation is extended to Ms. Marlene Linton and Ms. Debi Roof who typed the manuscript of this publication.

Robert E. Taylor
Executive Director
The National Center for Research
in Vocational Education

TABLE OF CONTENTS

	Page
FOREWORD	iii
LIST OF FIGURES	vii
SOME INTRODUCTORY REMARKS	1
How Does Your Career Education Instrument "Measure Up"?	2
PLANNING AND DESIGNING	7
What Are Some of the Problems Encountered When Attempting to Improve the Planning and Designing of Career Education Instruments	7
How Can I Improve the Planning and Designing of Career Education Instruments?	9
CHECKING RELIABILITY	19
How Can I Determine the Reliability of Career Education Instruments?	20
How Can I Improve the Reliability of Career Education Instruments?	26
DETERMINING VALIDITY	33
What Is Meant by the Term Validity?	33
How Can I Check a Career Education Instrument for Validity?	34
How Can I Improve the Validity of a Career Education Instrument?	36
ELIMINATING STEREOTYPES	41
Are My Career Education Instruments Free of Stereotypes?	41

Page

DEVISING THE FORMAT 43

How Can I Improve the Format of My Career Education Instrument? 43

GAUGING READABILITY 45

How Can I Improve the Readability of a Career Education Instrument? 45

EXAMINING YOUR CAREER EDUCATION INSTRUMENT 47

Determining Content Validity 47

Determining Internal Consistency Reliability 47

SOME CONCLUDING REMARKS 49

REFERENCES 51

LIST OF FIGURES

		Page
Figure 1.	A Checklist for Rating How Your Career Education Instrument "Measures Up"	3
Figure 2.	An Example of the Relationship Between Goals and Objectives	10
Figure 3.	A Planning Framework for Applying for a Job Test	12
Figure 4.	Cognitive Domain Classes and Example Career Education Assessment Activities	14
Figure 5.	Affective Domain Classes and Example Career Education Assessment Activities	15
Figure 6.	Psychomotor Domain Classes and Example Career Education Assessment Activities	16
Figure 7.	A Possible Planning Matrix for a Career Education Instrument on Applying for a Job	17
Figure 8.	Suggested Guidelines for Interpreting Reliability Coefficients	19
Figure 9.	Rank Order of Students on the Work Attitude Test	21
Figure 10.	Agreement Between Rankings on the Work Attitude Test	21
Figure 11.	Scores and Ranks for Individuals on a Test of Knowledge of the World of Work	23
Figure 12.	Agreement Between Rankings on the Knowledge of the World of Work Test	23
Figure 13.	Rank Order of Student Performances on Two Tests on Interpersonal Relationships	25
Figure 14.	Agreement Between Rankings on Two Tests on Interpersonal Relationships	25

Page

Figure 15. A Sample Distribution of Scores for a "Bell Shaped" Curve 27

Figure 16. A Comparison of How the Top and Bottom Groups Responded to Items in a Career Education Instrument 28

Figure 17. Two Sets of Directions for a Career Education Instrument 30

Figure 18. A Checklist for Improving Directions for Career Education Instruments 31

Figure 19. A Suggested Form for a Panel Review of an Economic Understanding Instrument 35

Figure 20. A Sample Form for Recording Test Items by Classes of the Cognitive Educational Domain 38

Figure 21. A Sample Form for Recording Test Items by Content Area 39

Figure 22. A Sample Form for Recording Scores When Determining Internal Consistency Reliability 48

SOME INTRODUCTORY REMARKS

As a career educator, you probably have been asked to review evaluation instruments that others have developed and you may have wanted to improve instruments that you developed. And, at times, you may have even questioned your ability to actually critique and improve these instruments. A number of questions come to mind when developing evaluation instruments that "measure up." How can you be sure that you have developed or adapted an effective evaluation instrument? How can you be confident of its effectiveness? How can you demonstrate the accuracy of its results? This handbook has been designed to assist you in answering these and other questions and thus improve the quality of your career education evaluation instruments.

You may ask, "Why is evaluation necessary?" Now, more than ever, evaluation is necessary to demonstrate an educator's accountability to the public. In 1976, the National Center for Educational Statistics estimated that educational institutions were expending more than $120 billion annually. Of this total, expenditures for elementary and secondary schools accounted for more than $75 billion. Naturally, people wonder what the schools are doing with these tax dollars. This is especially true if they assume (as reported by a 1976 Gallup Poll) that declining scores on standardized tests mean that the quality of education today is diminishing. Therefore, even with the widespread support for career education, educators such as yourself need to provide evidence of what students are achieving. In addition there is another factor that makes evaluation so important: evaluation should also provide information for improving the planning and development of career education programs. This type of information should enable you to do a better job of meeting the needs of students and to make sound educational decisions.

In order to demonstrate levels of student achievement, administrators and teachers have relied upon testing or measurement devices. In the case of career education, the lack of appropriate measures has, in many instances, caused individuals to develop their own instruments or to modify those developed by someone else. However, in most cases, administrators and teachers are not trained or experienced in test development and consequently lack confidence in the instruments that they have adopted.

The handbook has been designed to enable you to develop reliable instruments for evaluating your career education programs (at least before you consult with experts). This handbook provides basic information that should enable you to adapt and improve existing career education measurement devices. These improvement efforts have been organized around the following six topics: (1) Planning and Designing, (2) Checking Reliability, (3) Determining Validity, (4) Eliminating Stereotypes, (5) Devising the Format, and (6) Gauging Readability. A brief description of these topics is presented below.

- "Planning and Designing" explains how to structure revision efforts. The unit constitutes a "blueprint" for revising career education instruments.

- "Checking Reliability" reviews the important components of dependability and consistency in a career education instrument.

- "Determining Validity" poses the question, "Are you really measuring what you think you are measuring?" If you cannot tell what the measure is assessing, obviously the measure is of little value.

- "Eliminating Stereotypes" presents important ways to prevent the instrument from inhibiting responses because of an individual's sex or ethnic background.

- "Devising the Format" explains how the layout of the instrument can enhance the accuracy of the results from the instrument.

- "Gauging Readability" analyzes the ability of a career education instrument to communicate clearly with those for whom the device is intended. Often the "message" sent is not the "message" received.

The sequence of these six topics is not important. Each section presents central problems and some practical suggestions for minimizing them and improving the career education instrument. Simply select those of greatest relevance to your career education instrument. The following checklist has been developed for rating how your career education instrument "measures up" and selecting those topics that are most relevant for improvement efforts.

How Does Your Career Education Instrument "Measure Up"?

It is important to keep in mind that the higher the "quality" of a career education instrument, the more likely it is to capture information of which you can be confident. The lower the "quality" of a career education instrument, the less you can be confident of the information collected. Figure 1 presents a checklist you can use to assess your career education instruments. The checklist provides questions, response categories, and handbook references. The questions included on the checklist are keyed to the six major topics of the handbook. The handbook references cite the handbook pages containing the source material for the questions. The response column will enable you to rate your instrument for each of the questions asked.

In using the checklist, first read the question and mark your response in the appropriate space at the left of the question. If you are unsure of your response, refer to the column on the right and review the pages that are cited. Then, total your responses for each of the categories (i.e., Yes, No, Not Sure, or Not Applicable) and record each total in the space provided at the end of the checklist. Once you have totaled your responses, turn to page 6 to see what your responses mean.

Figure 1

A CHECKLIST FOR RATING HOW YOUR CAREER EDUCATION INSTRUMENT "MEASURES UP"

Response					Handbook Reference (Pages)
Yes	No	Not Sure	Not Applicable	Questions	
				PLANNING AND DESIGNING	
___	___	___	___	1. Did you review the goals and objectives for the career education topic to be assessed?	9 - 10
___	___	___	___	2. Did you develop an outline of the content areas to be included in your instrument?	11 - 12
___	___	___	___	3. Did you determine the classes of student behavior for the appropriate educational domain (e.g., cognitive, affective, psychomotor) that would be assessed by the instrument?	13 - 16
___	___	___	___	4. Did you check the relationship of each item to the classes of the appropriate educational domain and content areas to be assessed?	13, 17 - 18
				CHECKING RELIABILITY	
___	___	___	___	5. Did you determine the test-retest coefficient?	20 - 22
___	___	___	___	6. Did you determine the internal consistency coefficient?	22 - 24
___	___	___	___	7. Did you determine the equivalence coefficient?	24 - 26
___	___	___	___	8. Did you check for items that nearly everyone missed?	27
___	___	___	___	9. Did you check for items that nearly everyone got correct?	27
___	___	___	___	10. Did you check the correct and incorrect responses for the higher scoring individuals?	28 - 29
___	___	___	___	11. Did you check the correct and incorrect responses for the lower scoring individuals?	28 - 29
___	___	___	___	12. Did you consider the length of time available for testing?	29
___	___	___	___	13. Did you check the directions for clarity?	29
___	___	___	___	14. Did you check the directions for conciseness?	29
___	___	___	___	15. Did you avoid the use of ambiguous terms?	29 - 32
				DETERMINING VALIDITY	
___	___	___	___	16. Did you determine the content validity of the instrument?	34

Figure 1 (continued)

Response				Questions	Handbook Reference (Pages)
Yes	No	Not Sure	Not Applicable		
				DETERMINING VALIDITY (continued)	
—	—	—	—	17. Did you determine the concurrent validity of the instrument?	34
—	—	—	—	18. Did you determine the predictive validity of the instrument?	34, 36
—	—	—	—	19. Did you include examples of a wide variety of careers and occupations in the items?	36 - 37
—	—	—	—	20. Did you collect information about occupations to help improve the realism of the items?	36 - 37
—	—	—	—	21. Did you originally develop twice as many items as you intended to use in the instrument?	37
—	—	—	—	22. Did you have knowledgeable people review the instrument for relevance to the topic being assessed?	37
				ELIMINATING STEREOTYPES	
—	—	—	—	23. Did you include plural or both masculine and feminine pronouns in the items to avoid sexual stereotypes?	41
—	—	—	—	24. Did you use neutral terms in describing various occupations or careers?	41
—	—	—	—	25. Did you use names in the instrument that were different from Anglo in origin?	42
—	—	—	—	26. Did you include items that reflect a variety of life styles (e.g., single parent, families, extended families, suburban, urban, rural, etc.)?	42
—	—	—	—	27. Did you include items that reflect a variety of personalities (e.g., outgoing, shy, verbal, nonverbal, optimistic, pessimistic, confident, insecure, etc.)?	42
—	—	—	—	28. Did you include items that reflect a variety of socioeconomic levels?	42
—	—	—	—	29. Did you include items that reflect a variety of cultures?	42
—	—	—	—	30. Did you include some items that reflect individuals with handicapped conditions?	42
—	—	—	—	31. Did you include some items that reflect males in traditionally female roles and/or occupations?	42
—	—	—	—	32. Did you include some items that reflect females in traditionally male roles and/or occupations?	42

Figure 1 (continued)

Response				Questions	Handbook Reference (Pages)
Yes	No	Not Sure	Not Applicable		
				DEVISING THE FORMAT	
___	___	___	___	33. Did you limit the different items on your instrument to two or three types (e.g., multiple-choice, true-false, Likert, etc.)?	43
___	___	___	___	34. Did you use large type in preparing the instrument?	43
___	___	___	___	35. Did you use clear illustrations in preparing the instrument?	43
___	___	___	___	36. Did you keep all possible responses to an item on the same page as the question?	43 - 44
___	___	___	___	37. Did you prepare an attractive page layout?	44
				GAUGING READABILITY	
___	___	___	___	38. Did you begin with relatively easy items?	45
___	___	___	___	39. Did you test the instrument for readability?	45
___	___	___	___	40. Did you pilot test the instrument with individuals similar to those for whom the measure was intended?	45
___	___	___	___	**TOTALS**	

A high total of "yes" responses suggests that you have a "good" instrument. Just for fun, let's determine a "grade" for your instrument; the following grading scale is based upon the total number of "yes" responses that you obtain:

37 or more = A (Move to the top of the class)

33 to 36 = B (Better than most)

30 to 32 = C (Not bad—but there is room for improvement)

26 to 29 = D (You had better reconsider)

25 or less = F (Have you thought about starting over?)

PLANNING AND DESIGNING

Planning provides you with an opportunity to carefully examine the purpose, content and use of a career education instrument. Yet, as a general rule, planning receives insufficient attention when an instrument is being revised. Prior to revising an instrument, you need to have a clear picture of what you wish to assess. Take a few minutes to reflect on your instrument, describe the purpose of the instrument in as direct a manner as possible; think about the items (questions) that you have included; consider the relationship of the individual items to the instrument's purpose. Do all of your items relate to the purpose of the instrument? You may find it necessary to eliminate or revise some items that you had originally included or to add additional items. This section of the handbook will discuss some of the common problems encountered in planning career education instruments and some practical ways for revising these instruments.

What Are Some of the Problems Encountered When Attempting to Improve the Planning and Designing of Career Education Instruments?

There are a number of problems that impede the planning and designing of career education instruments. First, the real purpose of the instrument is often not clearly established. You need to question the reason for the instrument and the items it contains:

- Is the instrument designed to assess students' specific knowledge?
- Is the instrument designed to assess students' ability to make judgments?
- Is the instrument designed to determine students' ability to perform certain tasks or to achieve certain skills?
- Is the instrument designed to assess some combination of the skills listed above or even to serve other purposes?

Career education instruments often test trivial matters when insufficient time has been devoted to developing the items on the test. Items that ask simply for recall of certain facts or figures are generally easier to develop than are those that test for understanding or ability to apply specific knowledge.

Items are frequently borrowed from previously developed tests. Utilizing existing items is easier than developing new items, especially when you are under pressure to develop a test for immediate use. However, when items are taken from existing tests, it is often difficult to relate them sufficiently to the goals, outcomes, or elements of your career education program.

We are often unaware of the various uses that can be made of tests. Katz (1973) identified eight different uses of test scores. You may have other uses that you would like your instrument to serve. Examples of Katz's uses of test scores as they apply to career education are presented on the following page.

Placing students — More and more emphasis is being placed upon open-entry and open-exit types of educational activities. In career education, there is a need to consider where a student is, in relationship to a particular activity rather than assume that all students must start at the same point and spend the same amount of time on a given activity. For example, a student with previous work experience may not need the same amount of information on job-seeking skills as one who has not had work experience. Therefore, one possible use of a career education instrument on job-seeking skills could be to place a student in a group requiring similar information.

Diagnosing pupils' weaknesses — It is also possible to use tests to identify specific strengths and weaknesses of individuals. Let us assume that you want to determine the particular aspects of work attitudes with which students need assistance. After administering a test of work attitudes, you might find that some students have a high understanding of the need to be loyal to their employers and have positive attitudes toward putting in a day's work for a day's pay. While other students have a low understanding of the importance of being consistent and persistent. In this case, you could use an individual's score on various subparts of a work attitude test in prescribing activities to increase identified deficiencies.

Assessing effectiveness of instruction — A question often asked of an educational program is, does it work? People are often reluctant to adopt or adapt new practices without evidence as to their effectiveness. In this case, a test could be used to compare individuals' occupational knowledge prior to career education activities with their self-concept after the activities. The resulting difference could then be used as an indication of the effectiveness of career education.

Predicting pupils' achievement — A test might also be used to predict pupils' achievement at some later point in life. For example, some might argue that students who are knowledgeable of the variety of potential careers available to them make better career choices, and, subsequently, are more satisfied with their employment. In this case, students' scores on a test of career opportunities could be compared with a test of their job satisfaction at some later point in time. The scores could then be compared to determine the accuracy with which students' scores on a career awareness test did, in fact, predict their satisfaction with their employment.

Evaluating schools — In many districts, student achievement scores are routinely compared by schools and buildings. This allows administrators to see how various schools are doing both within the district and in comparison to other districts such as theirs. Often this type of information is used in releases to the school board as well as to the general public. A comparison of average scores by school, on a career development test, would be an example of this type of test usage.

Comparing a pupil's achievement to a fixed standard — The public is becoming more and more concerned that students are unable to perform what they consider "basic life skills." Although many arguments (both pro and con) can be presented, it appears that some schools are establishing minimal levels of students' knowledge, attitudes, and skills. For example, a person could be required to complete a job application blank prior to being advanced or graduated. In this case, a student could be given an application form and asked to complete it, or

respond to specific questions related to application forms. In either case, a minimum score could be established and students would be required to perform at this level prior to advancement or graduation.

Comparing a pupil's achievement to a score on another test — A student's time is one of the limited resources with which educators must work. Consequently, this time must be used as efficiently and effectively as possible. In career education, we should not spend time obtaining evaluation information that might already be available. Therefore, test scores of the various aspects of career education should be compared with other test scores to see if there are any tests that might be eliminated. To illustrate, if scores on a career planning test are highly correlated with scores on a career information test already in use, it may be possible to drop one of the tests from the testing and evaluation program.

Comparing a pupil's achievement to a teacher's marks — A final use that can be made of test scores is to compare the results of an achievement test with the marks of a teacher. In many cases, teacher's marks are thought to reflect many interpersonal characteristics of the student and are not truly indicative of knowledge, attitude, or skills of the student. Therefore, an achievement test is often thought to be a more objective estimate of a pupil's achievement than are a teacher's marks. In career education, for example, an assessment of work attitudes via an independently developed test would generally be viewed more credibly than would teacher rankings of students' work attitudes.

We have noted a few of the complications you might encounter in planning and revising instruments. So to forestall these problems, we will now discuss some practical ways of improving the way that career education instruments are planned and revised.

How Can I Improve the Planning and Designing of Career Education Instruments?

Several techniques are available to those who wish to improve the planning and designing of career education instruments. This section develops the following four suggestions:

1. Before revising the instrument, review information on the goals and objectives of the career education topic to be assessed.
2. Develop a basic outline of what is to be included in the instrument.
3. Determine the types of student behavior associated with the area to be tested.
4. Develop a planning matrix.

1. Reviewing information on the goals and objectives. In order to improve the quality of an instrument, you should review the goals and objectives of the career education topic to be assessed. Despite the fact that research in career education is complicated by the absence of definitions, you should review the goals, elements, and objectives that shaped your career education program. In addition to your own goals and objectives, you may want to review the goals, elements, and outcomes

of other career education programs.* Other resources for researching the topic you wish to assess range from the Educational Research Information Center (ERIC) to the card catalog in your local library. And don't hesitate to consult the numerous authorities available, such as your state career education director or personnel from planning, research, and evaluation units. Other knowledgeable individuals are located in colleges and universities, educational laboratories and centers, and private agencies. Larger school districts may also have planning, research, and evaluation units that would be helpful to you in this process.

But remember your program has its own goals and objectives and these must be kept foremost as you revise your career education instruments. As you review the goals and objectives of the career education topic you wish to assess, it will be helpful to prepare a chart that illustrates the relationships of the objectives to the goals. A sample of this type of chart is presented in Figure 2.

Figure 2

AN EXAMPLE OF THE RELATIONSHIP BETWEEN GOALS AND OBJECTIVES

Sample Goals of a Career Education Program	Sample Objectives of a Career Education Program
1. To increase the ability of students to obtain employment.	1. Students will be able to locate potential employers. 2. Students will be able to neatly and accurately complete an application for employment. 3. Students will demonstrate desirable behaviors during an interview situation.
2. To understand how one makes career related decisions.	1. Students will be able to identify situations with career related problems. 2. Students will be able to list the steps followed in solving career related problems. 3. Students will distinguish between the causes and symptoms of career related problems.

**Developmental Program Goals for the Comprehensive Career Education Model* (1972); *Michigan Career Education: Career Development Goals and Performance Indicators* (1974); *Basic Learner Outcomes for Career Education* (1973); *Perspectives on the Problem of Evaluation in Career Education* (1976); *Evaluation and Educational Decision Making: A Functional Guide to Evaluating Career Education* (1975).

2. Developing a content outline. As mentioned previously, it is essential that you have a clear picture of what you are trying to assess before you begin work on the instrument. An outline will help you to clarify your purpose.

Let us suppose that you are developing a test on "applying for a job." One of the first determinations you should make is whether this topic is logically connected to the goals, outcomes, or elements of your entire project. For example, if one of your goals is to help students obtain employment, the logical relationship of "applying for a job" to the goal is quite clear (see Figure 2). Often, however, there are discrepancies between topics and goals that should be detected before imprecise items are devised. Furthermore, relating your topic to your goal will help you to generate additional information or a broadened outline. For example, applying for a job is one of four skills related to obtaining employment that the state of Florida (1973) identified in the following portion of an outline:

- Obtaining Employment
  - A. Choosing an occupation
  - B. Locating potential employers
  - C. Securing initial information about opportunities
  - D. Applying for a job

You may agree or disagree with this conceptualization of obtaining employment; however, the important point is that there should be a test of the logical connection between the topic and the goals or outcomes for the career education program.

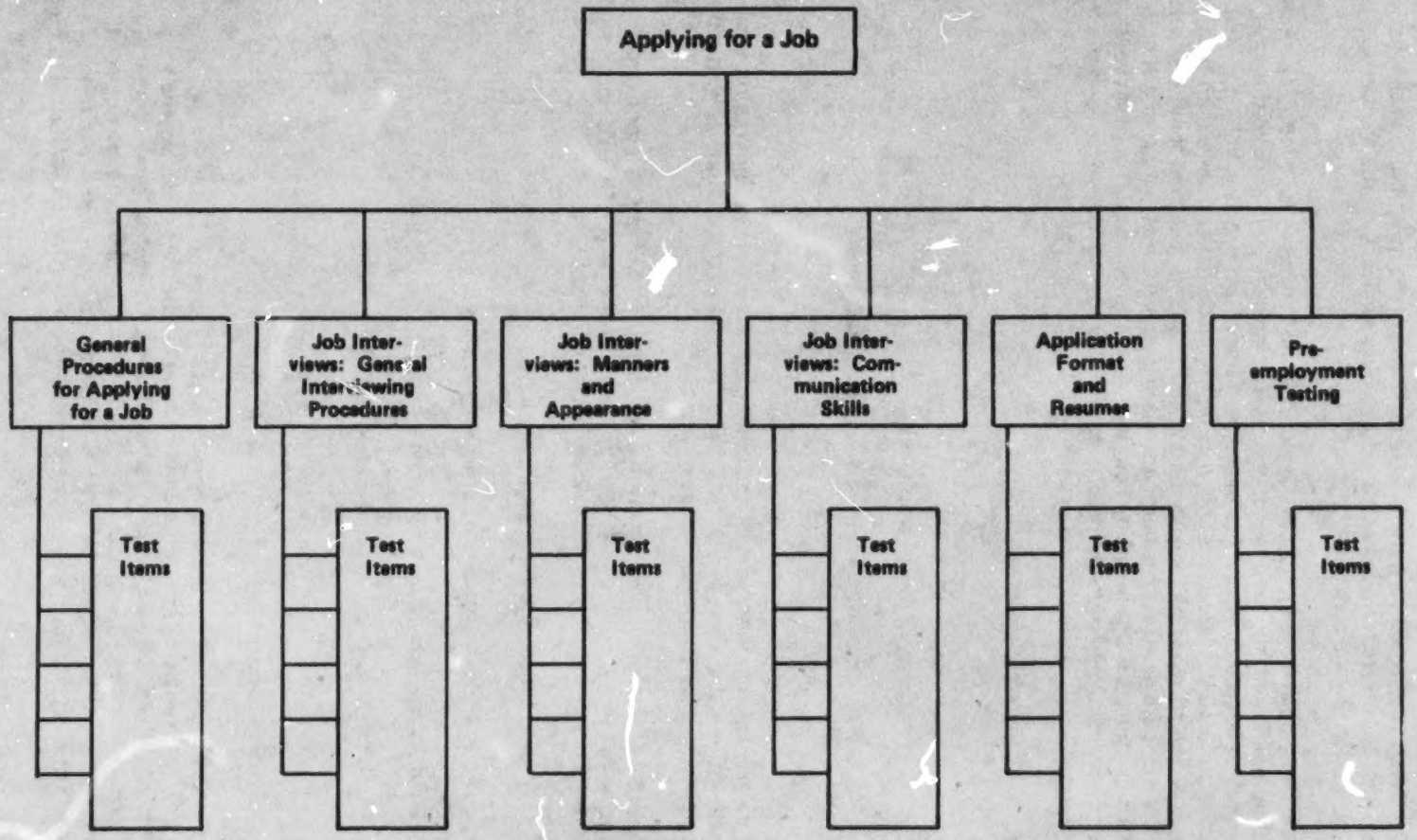
Another step in generating the outline is to consider subheadings; for example, what activities are involved in "applying for a job"? Again, the state of Florida (1973) determined that the topic of applying for a job consists of the following content areas:

- General procedures for applying for a job
- Job interviews: general interviewing procedures
- Job interviews: manners and appearance
- Job interviews: communication skills
- Application forms and resumes
- Preemployment testing

These content areas can then be used as major sections, around which the individual items can be developed. An example of how these content areas can be used in developing a planning framework for a test on applying for a job is presented in Figure 3. If each of these content areas is of equal importance, the number of items for each section should be the same. If some of the content areas are more important than others then they should contain more items than those of lesser importance. These content areas also form one dimension of the planning matrix presented in Figure 7.

Figure 3

A PLANNING FRAMEWORK FOR TEST ON APPLYING FOR A JOB TEST



3. Relating student behavior. Bloom (1956) classified student behavior into three domains: *cognitive, affective,* and *psychomotor.* Recalling these domains will help you to state precisely and realistically the desired outcomes for students. After these domains are surveyed, they will be combined with the content areas listed above to form a matrix of factors to consider when planning career education instruments.

The *cognitive* domain includes skills that require the recall or recognition of knowledge and the development of intellectual abilities. The cognitive domain contains the following six major classes: knowledge, comprehension, application, analysis, synthesis, and evaluation. These classes are arranged in order from the simple to the complex. Definitions of each of these classes and examples of career education assessment activities are presented in Figure 4. For further information related to this domain, you may want to review the *Taxonomy of Educational Objectives: The Classification of Educational Goals. Handbook I: Cognitive Domain* (Bloom, et al., 1956).

The *affective* domain includes behavior that reflects an individual's feelings, emotions, or degree of acceptance or rejection. The affective domain includes the following five classes: receiving, responding, valuing, organizing, and characterizing by a value or value complex. Again, these classes are arranged from the simple to the complex. Definitions of each of these classes and examples of career education assessment activities are presented in Figure 5. For further and more detailed information on the affective domain, you may want to study the *Taxonomy of Educational Objectives: The Classification of Educational Goals. Handbook II: Affective Domain* (Krathwohl, et al., 1956).

The *psychomotor* domain includes actions that require some muscular or motor skill, some manipulation of material and objects, or some act which requires neuromuscular coordination. The psychomotor domain includes the following four classes: gross bodily movements, finely coordinated movements, nonverbal communication behaviors, and speech behaviors. For additional information on this domain, you are encouraged to review *Behavioral Objectives and Instruction* (Kibler, et al., 1970).

Although these three domains have been presented as mutually exclusive categories, determining the proper classification of a given behavior is not always a simple task. Furthermore, the classes within each domain often overlap, making it even more difficult to make the proper classification. Therefore, when trying to classify items according to educational domains and classes you will probably have to make some arbitrary decisions regarding their "best fit."

4. Utilizing a planning matrix. You can develop a matrix that will help you examine the composition and number of items for each topic to be included in an evaluation instrument. In such a matrix, the classes of the appropriate domain are placed along the left hand side of the matrix and the content areas are placed across the top of the matrix. It should be noted that it would be quite unusual to include measurement related to all three (cognitive, affective, and psychomotor) of the domains in the same instrument. As a result of using a planning matrix, you can revise your career education instruments to more adequately assess the aspects of career education that you deem to be most important. Figure 7 presents a matrix that was developed to examine an instrument on "applying for a job."

The six content areas included in "applying for a job" that were identified earlier are placed across the top of the matrix. Similarly, the six classes of the cognitive domain that we discussed earlier are placed down the side of the matrix. As a result, the matrix contains a total of 36 different combinations of the classes of the cognitive domain of educational behavior and the content areas included in "applying for a job." Each of these combinations represents one cell in our

Figure 4

COGNITIVE DOMAIN CLASSES AND EXAMPLE CAREER EDUCATION ASSESSMENT ACTIVITIES

Classes	Examples of Career Education Assessment Activities
Knowledge—Remembering, either by recognition or recall, of ideas, material, or phenomena.	Ask students to identify tools or equipment that workers in a particular occupation would be expected to use.
Comprehension—Understanding what is being communicated and being able to make some use of the material or ideas contained in it.	After reading about an occupation, have the students respond to a set of questions about the occupation or prepare a short essay on their ability to work in the occupation and interest in the occupation.
Application—Requires the act of problem solving, using acquired knowledges and understandings. The problems should be new and different from the situations in which material was originally presented.	Provide an example of a personal conflict in a work setting and ask the students how they would minimize the conflict.
Analysis—The ability to break down material into component parts.	Provide a scenario of someone selecting a new job and ask the students to distinguish relevant arguments for the job from irrelevant arguments for the job.
Synthesis—Putting elements and parts together so as to form a whole.	Ask students to prepare a written response to a letter in which an individual complains about unsatisfactory service.
Evaluation—Making judgements about the value, for some purpose of ideas, works solutions, methods, material, etc.	Ask the students to rank potential applicants for a particular job based upon: (1) a completed application and (2) vita.

Figure 5

AFFECTIVE DOMAIN CLASSES AND EXAMPLE CAREER EDUCATION ASSESSMENT ACTIVITIES

Classes	Examples of Career Education Assessment Activities
Receiving—Deals with a learner's willingness to learn.	Ask students to indicate likes and dislikes related to various careers or willingness to study about careers.
Responding—Is concerned with learner's level of commitment to active participation.	Ask students about their feeling of satisfaction as it relates to various types of work that they have done.
Valuing—Emphasizes the development of attitudes, beliefs, and values as it relates to the worth of a thing, phenomenon, or behavior.	Ask students to indicate whether they agree or disagree with statements regarding various work habits.
Organizing—Requires the learner to bring together values to form a value system.	Describe a life style and four possible attitudes related to the life style. Ask students to arrange the attitudes in order of preference.
Characterizing by a Value or Value Complex—Refers to an organization of attitudes and values that have been internalized into a consistent system or philosophy which guides the individual.	Provide students with a list of different occupations. Next, pair each occupation with each of the remaining occupations on the list. Then ask students to indicate which one in the paired occupations on the list is most indicative of the type of work they would like to do.

Figure 6

PSYCHOMOTOR DOMAIN CLASSES AND EXAMPLE CAREER EDUCATION ASSESSMENT ACTIVITIES

Classes	Examples of Career Education Assessment Activities
Gross Bodily Movement—Is characterized by movement of entire limbs.	Explain to students that as workers in health careers they may be required to move patients from one area to another. To assess strength, ask students to move an object or weight from one point to another.
Finely Coordinated Movements—Involves movement of the extremities, e.g., hands, fingers, feet, etc.	Give students handwritten letter and ask them to type the letters within a given time frame.
Nonverbal Communication Behaviors—Centers on facial expressions, gestures, and/or bodily movements.	Set up an interview setting, and ask students to demonstrate enthusiasm without actually speaking.
Speech Behaviors—Relates to an individual's ability to verbally express himself/herself.	Describe a work situation and ask students to "state" the consequences of poor work habits.

Figure 7

A POSSIBLE PLANNING MATRIX FOR A CAREER EDUCATION INSTRUMENT ON APPLYING FOR A JOB

Classes of the Cognitive Domain	Content Area							
	General Procedures for Applying for a Job	Job Interviews: General Interviewing Procedures	Job Interviews: Manners and Appearance	Job Interviews: Communication Skills	Application Forms and Resumes	Pre-employment Testing	Total Number of Items	Percent of Total Items
Knowledge	1	1	2	1	1	1	7	14
Comprehension	1	1	2	1		4	9	18
Application		1		1	5		7	14
Analysis		3	3	3	5		14	28
Synthesis	5	1		1		1	8	16
Evaluation	1	1	1	1	1		5	10
Total Number of Items	8	8	8	8	12	6	50	100
Percent of Total Items	16	16	16	16	24	12	100	

planning matrix. In examining a test, each item is read and classified according to the class of educational domain and content area that it represents. Once this determination has been made, a dash mark is made in the appropriate cell of the planning matrix. Once the examination of the test is completed, the number and percent of items for each class and content area can be computed.

It should also be noted that not all of the cells in the matrix will necessarily have entries. Figure 7 indicates that a total of 50 items were included in the test; that number appears in the total-number-of-items cell for the matrix. Look again at Figure 7, you will note that the most important content area is application forms and resumes; and preemployment testing is the least important content area. Similarly, the analysis class is the most important and evaluation is the least important class of the cognitive domain.

By using a planning matrix in revising an instrument, you will be able to determine the relative emphasis for each of the classes of domains and content areas included in your test. Without using a planning technique such as a matrix, you could unintentionally place emphasis on certain classes or content areas and thus develop an instrument that does not correctly measure what is intended. This is often the case when instruments are developed utilizing only items that are easy to write and score.

CHECKING RELIABILITY

In describing a person as reliable, you might say the person is trustworthy, dependable, stable, and consistent; and you would be confident of the person's ability to perform. The term "reliability" as applied to an evaluation instrument has much the same meaning. In other words, a reliable instrument will provide similar results each time it is used. It is important that you be confident of the results obtained with evaluation instruments you use. Otherwise, it will be impossible for you to rely on the results. Therefore, you need to determine the reliability of your test prior to placing too much confidence in your results.

Reliability estimates can be established according to the different formulas found in such basic measurement texts as Cronbach's *Essentials of Psychological Testing* (1970) and Ebel's *Measuring Educational Achievement* (1965). As a general rule, reliability is expressed as a decimal number (coefficient) ranging from 0.00 to +1.00. The closer the decimal (coefficient) is to +1.00, the more reliable the test is said to be. Figure 8 suggests general guidelines to follow in interpreting the adequacy of reliability coefficients.

Figure 8

SUGGESTED GUIDELINES FOR INTERPRETING RELIABILITY COEFFICIENTS

Reliability Coefficients		Adequacy
.90 and higher	=	Super
.80 – .89	=	Not bad
.70 – .79	=	OK, but . . .
.60 – .69	=	Better watch out
Less than .60	=	No way!

Katz (1973) indicated that there are three basic kinds of reliability: stability, internal consistency, and equivalence. Each of these types will be discussed in this section.

Stability (or test-retest reliability) is the likelihood that an instrument will elicit the same levels of performance at different testing times. Test-retest reliability is determined by comparing pretest scores with postscores.

Internal consistency is the similarity of scores on different parts of a test that are taken at the same time. As a general rule, internal consistency is determined by correlating the score on odd-numbered items with the score on even-numbered items.

Equivalence is the correlation between scores on different testing forms of the same instrument. Equivalence is determined simply by comparing scores on two or more separate forms of the same instrument.

The purpose of this section is not to develop the capability to compute complex reliability indices for instruments; rather, it is to show how easily locally developed instruments can be improved through paying attention to reliability issues.

How Can I Determine the Reliability of Career Education Instruments?

There are a number of reference works that analyze in-depth the statistical and mathematical theory involved in determining reliability. This section, however, is designed to provide more simplified methods for determining the three kinds of reliability.

Stability (test-retest) is probably the easiest kind of reliability to determine because it only requires administering an instrument two times to the same group of students and then comparing the scores. Nevertheless, several problems need to be recognized. The most obvious one is that the students tend to remember the questions and answers from one administration of the instrument to the next. Another problem is that students often feel that taking a test twice is a waste of their time; this attitude alone can lessen the reliability of the results. Still another problem is that if the time between the two administrations is too long, the students' responses may differ because of learning that occurred between the two administrations.

One way to determine the test-retest reliability could be to administer your instrument immediately prior to and after a school vacation lasting for one or two weeks. This would eliminate the effect of any instruction that might occur between the two administrations of the instrument.

Recognizing the fact that these problems exist, let's consider the following situation. Let's assume that you have developed an instrument to assess students' work attitudes. Let's also assume that, in an attempt to eliminate (or minimize) the effect of learning that could occur between administrations of the test, you administered the test immediately before and immediately after spring vacation.

You will recall from page 19 that the closer the reliability coefficient is to +1.00, the more reliable the instrument is considered to be. You can compute the "stability coefficient" of your instrument using the Kendall rank correlation coefficient T (tau). Siegel (1956) suggests the following formula for tau:

$$T = \frac{S}{\frac{1}{2}N\,(N-1)}$$

In Siegel's formula, S = the number of individuals on the second administration of the work attitude test that are in the same order as they were on the first administration of the work attitude test, and N = the number of students which were ranked on both administrations of the work attitude test. To determine the value of tau, start by arranging the students' names in sequential order

from the highest score to the lowest score for each of the two separate administrations of the test (as shown in Figure 9).

Figure 9

RANK ORDER OF STUDENTS ON THE WORK ATTITUDE TEST

Result of First Administration	Result of Second Administration
1. Marlene Lee	1. Marlene Lee
2. Trenton Bayes	2. Jose Gonzales
3. Kay Angona	3. Kay Angona
4. Jose Gonzales	4. Trenton Bayes
5. Craig Garrett	5. Kelli Hall
6. Jerry Wong	6. Pauline Jacobinski
7. Kelli Hall	7. Craig Garrett
8. Pauline Jacobinski	8. Jerry Wong
9. Barry Wilcox	9. Barry Wilcox
10. Richard Alston	10. Jack Vance
11. Martin Stone	11. Jeanette Lang
12. Jack Vance	12. Martin Stone
13. Jeanette Lang	13. Mary Ching
14. Mary Ching	14. Steve Nelson
15. Steve Nelson	15. Richard Alston

Next, compile a table like the one shown in Figure 10. Examine Figure 10, note that there are three columns. The first column lists the names of students in order (from highest to lowest) according to their scores on the first administration of the test. Column 2 lists their respective rank on the first administration of the test and column 3 lists their rank on the second administration of the test.

Figure 10

AGREEMENT BETWEEN RANKINGS ON THE WORK ATTITUDE TEST

Column 1 Individual	Column 2 Rank on First Administration	Column 3 Rank on Second Administration
Marlene Lee	1	1
Trenton Bayes	2	4
Kay Angona	3	3
Jose Gonzales	4	2
Craig Garrett	5	7
Jerry Wong	6	8
Kelli Hall	7	5
Pauline Jacobinski	8	6
Barry Wilcox	9	9
Richard Alston	10	15
Martin Stone	11	12
Jack Vance	12	10
Jeanette Long	13	11
Mary Ching	14	13
Steve Nelson	15	14

You are now ready to compute S, which in turn will enable you to determine the degree of agreement between the scores received on the two administrations of the test.

To find the value of S, start with the rank of the first student on the second administration of the test and count the number of ranks below it which are larger; then subtract from this number the number of ranks below it that are smaller. After completing this process for all ranks, sum the results to obtain S. Thus, in our example the rank which is listed first on the second administration is 1. This rank has 14 ranks which are larger below it and 0 ranks which are smaller. Therefore, its contribution to S is (14-0). The next rank is 4. It has 11 ranks below it that are larger and 2 ranks that are smaller so that the contribution to S is (11-2). By proceeding in this manner you can see that:

$$S = (14\text{-}0) + (11\text{-}2) + (11\text{-}1) + (11\text{-}0) + (8\text{-}2) + (7\text{-}2) + (8\text{-}0) + (7\text{-}0) + (6\text{-}0) + (0\text{-}5) + (2\text{-}2) + (3\text{-}0) + (2\text{-}0) + (1\text{-}0) = 77$$

Knowing that 15 students were administered the work attitude test, we can proceed to compute the stability coefficient:

$$T = \frac{77}{\frac{1}{2}(15)\ (15\text{-}1)}$$

$$T = \frac{77}{105}$$

$$T = .73$$

In this case the stability coefficient is only .73. According to Figure 8, this test is "OK, but. . . ." Before placing complete confidence in this test it should be made more reliable. Pages 26-32 present ways to improve reliability.

Internal consistency (or split-half reliability) can be determined by performing the following three steps. First, divide the test into two parts (e.g., on the basis of odd- and even-numbered items). Then, score each part separately, as if it were the entire test. Finally, compute an internal consistency coefficient to determine the degree of agreement between the two scores.

Consider this example: let us assume that you have developed a fifty-item test gauging knowledge of the world of work. Divide the test according to odd- and even-numbered items into two halves of twenty-five items each. Score each half as you would any other test. Then (as in Figure 11) record the two scores for each individual and determine the rank order of these individuals.

In computing the internal consistency coefficient you can use the same basic procedure that was suggested for determining the stability coefficient. First, arrange the names in rank order of their scores on the even-numbered items. Second, place the rank order of each student on the even-numbered items in a column to the right of the name. These numbers should be in sequential order from one to the number of students that took the test (in this case 15). Third, place the rank order of each student on the odd-numbered items in a second column to the right of the first column. An example of this procedure is presented in Figure 12.

Again, the procedure suggested by Siegel (1956) can be used to compute Kendall's rank correlation coefficient (tau) for the internal consistency coefficient:

Figure 11

SCORES AND RANKS FOR INDIVIDUALS ON
A TEST OF KNOWLEDGE OF THE WORLD OF WORK

Individual	Score on Even Half	Student Rank on the Even Half	Score on Odd Half	Student Rank on the Odd Half
Richard Alston	72	14	75	14
Kay Angona	89	5	90	5
Trenton Bayes	63	15	65	15
Mary Ching	90	4	88	6
Craig Garrett	86	8	86	8
Jose Gonzales	91	3	92	4
Kelli Hall	83	11	83	11
Pauline Jacobinski	92	2	93	3
Jeanette Lang	82	12	94	2
Marlene Lee	93	1	95	1
Steve Nelson	85	9	82	12
Martin Stone	80	13	81	13
Jack Vance	87	7	85	9
Barry Wilcox	88	6	84	10
Jerry Wong	84	10	87	7

Figure 12

AGREEMENT BETWEEN RANKINGS ON
THE KNOWLEDGE OF THE WORLD OF WORK TEST

Column Name	Column 2 Rank on Even-Numbered Items	Column 3 Rank on Odd-Numbered Items
Marlene Lee	1	1
Pauline Jacobinski	2	3
Jose Gonzales	3	4
Mary Ching	4	6
Kay Angona	5	5
Barry Wilcox	6	10
Jack Vance	7	9
Craig Garrett	8	8
Steve Nelson	9	12
Jerry Wong	10	7
Kelli Hall	11	11
Jeanette Lang	12	2
Martin Stone	13	13
Richard Alston	14	14
Trenton Bayer	15	15

$$T = \frac{S}{\frac{1}{2} N \ (N-1)}$$

In this analysis, S = the number of individuals on the scoring of the odd-numbered items that are in the same order as their score on the even-numbered items and N = the total number of students who took the knowledge of the world of work test. In determining the value of S, start with the rank of the first student on scoring of the odd-numbered items and count the number of ranks below it that are larger. Then, subtract from this number the number of ranks below it that are smaller. Do this for each of the entries in Column 3 and sum the totals to obtain S. To illustrate, the rank of the first student of the odd-numbered scoring is 1. Below this rank are fourteen ranks that are larger and none that are smaller. Therefore, the component of the S value for this rank would be 14-0 or 14. Similarly for the second student, there are twelve ranks larger than 3 and one that is smaller than 3. Hence, the component for this entry would be 12-1 or 11. Continue this process and you can see that:

$$S = (14-0) + (12-1) + (11-1) + (9-2) + (9-1) + (5-4) + (5-3) + (5-2) + (3-3) + (4-1) + (3-1) + (3-0) + (2-0) + (1-0) = 67$$

Again, since fifteen students were administered the knowledge of the world of work test, you can compute the internal consistency coefficient by substituting this value for N:

$$T = \frac{67}{\frac{1}{2} (15) \ (15-1)}$$

$$T = \frac{67}{105}$$

$$T = .64$$

Refer again to Figure 8; you will note that this coefficient is marginal and the internal consistency of the test should be questioned. Therefore, you should not place much confidence in the results of this test. Pages 26-32 present methods for improving internal consistency.

Equivalence is determined by comparing an instrument's results with the results of a different test of the same material. Let us assume that two different teachers have each prepared a test designed to measure interpersonal relationships of ninth grade students, and you are interested in determining whether both instruments provide similar results. To measure the equivalence, you would first administer both instruments to the same group of students. Then you would rank order the students' names according to their scores on each test, with the highest scoring student listed first and the lowest scoring student listed last (as in Figure 13).

Once the rank order for each student on each of the two tests has been determined, you are ready to determine the agreement between the rankings on the two tests. To help in making this determination, compile a table like that shown in Figure 14. Arrange the names in rank order of their scores on Test A and list the names in Column 1. Place the number of the rank order in Column 2. These ranks should be in sequential order from one to the number of students who took the test (in this case ten). Next, place the rank order of each student on Test B in Column 3.

You are now ready to compute the equivalence coefficient following the procedure suggested by Siegel (1956) for the Kendall rank correlation coefficient or tau (T) where:

Figure 13

RANK ORDER OF STUDENT PERFORMANCES ON TWO TESTS ON INTERPERSONAL RELATIONSHIPS

Interpersonal Relationships Test A	Interpersonal Relationships Test B
1. Pauline Jacobinski	1. Martin Stone
2. Martin Stone	2. Mary Ching
3. Mary Ching	3. Pauline Jacobinski
4. Steve Nelson	4. Steve Nelson
5. Jose Gonzales	5. Kelli Hall
6. Kelli Hall	6. Trenton Bayes
7. Trenton Bayes	7. Jose Gonzales
8. Marlene Lee	8. Jerry Wong
9. Jerry Wong	9. Marlene Lee
10. Craig Garrett	10. Craig Garrett

Figure 14

AGREEMENT BETWEEN RANKINGS ON TWO TESTS ON INTERPERSONAL RELATIONSHIPS

Column 1 Individual	Column 2 Rank Order on Test A	Column 3 Rank Order on Test B
1. Pauline Jacobinski	1	3
2. Martin Stone	2	1
3. Mary Ching	3	2
4. Steve Nelson	4	4
5. Jose Gonzales	5	7
6. Kelli Hall	6	5
7. Trenton Bayes	7	6
8. Marlene Lee	8	9
9. Jerry Wong	9	8
10. Craig Garrett	10	10

$$T = \frac{S}{\frac{1}{2} N (N-1)}$$

As was the case in our two similar analyses, S = the number of individuals on the scoring of Test B that are in the same order as their scores on Test A and N = the total number of students who took both Test A and Test B. In computing the value of S, we take the rank of the first student on Test B and determine the number of ranks below it that are larger and subtract from that number the number of ranks below it that are smaller. Similarly, we proceed through the rankings of remaining students on Test B. Hence, you can see the following computation:

$$S = (7\text{-}2) + (8\text{-}0) + (7\text{-}0) + (6\text{-}0) + (3\text{-}2) + (4\text{-}0) + (3\text{-}0) + (1\text{-}1) + (1\text{-}0) = 35$$

In this case, the total number of students taking both Test A and B was ten. Therefore, we can substitute into our formula accordingly and see that:

$$T = \frac{35}{\frac{1}{2} (10) (9)}$$

$$T = \frac{35}{45}$$

$$T = .78$$

In this case, we should be cautious (see Figure 8) in stating that these tests are equivalent. Each test should be carefully examined to determine which test more accurately reflects the desired measurement. The following section presents ways to improve the reliability of evaluation instruments that you develop.

In this section we have discussed the overall reliability of your career education instruments. As a result of studying this section, you should be able to determine if the reliability of your career education instrument needs to be improved. The section that follows indicates what you can do with specific sections and items to improve the reliability of your career education test.

How Can I Improve the Reliability of Career Education Instruments?

The improvement of an instrument's realibility is one evaluation activity from which career education personnel can receive immediate and direct benefits without having to rely upon other educational agencies for materials. This section presents six practical suggestions for improving the reliability of your career education instruments:

1. Remove the items from the instrument that everyone misses and those that everyone gets correct.

2. Remove the responses that no one selects and those that everyone selects.

3. Retain items on which the higher scoring individuals were successful and the lower scoring individuals were unsuccessful.

4. Increase the length of the test.

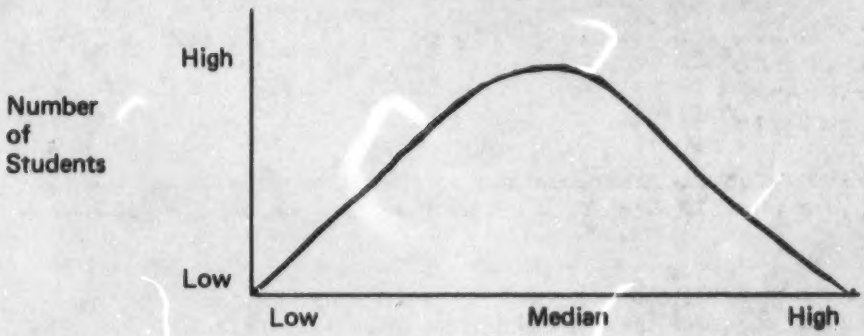
5. Provide clear and concise directions to the individual taking the test.

6. Eliminate ambiguous items.

The first three suggestions listed above are particularly applicable to career education instruments that are designed to discriminate highly knowledgeable individuals from less knowledgeable individuals. Tests of this type are called norm-referenced. Norm-referenced testing compares individual performance of a student against the performance of a particular group of students. Responses to a norm-referenced test are generally distributed in a normal "bell shaped" curve. An example of the "bell shaped" curve is shown in Figure 15.

Figure 15

A SAMPLE DISTRIBUTION OF SCORES FOR A "BELL SHAPED" CURVE



From a test development point of view, you would not want to include items which the students with low knowledge levels get correct and the students with the high knowledge levels get incorrect. The remainder of this section will briefly discuss each of the six suggestions mentioned previously.

1. **Removing items that everyone answers correctly or answers incorrectly.** A reliable instrument will identify persons who have achieved expected career education outcomes, so if an item on the instrument is answered the same way by everyone, the item does not help to distinguish performances and thus fails to contribute to the instrument's reliability. The process of isolating these items is simple. Review the instrument item by item, recording for each item, the number of persons who answered it incorrectly. The person scoring the instrument could record the number directly on the instrument. Or the teacher could ask the students to indicate (by a show of hands) who got the item right or wrong; furthermore, discussing the reasons for the various responses will elicit information that could then be used to develop more reliable items.

2. **Removing the responses that no one or everyone selects.** As described above, any item that does not help to distinguish persons who have achieved an outcome from those who have not should be eliminated from the instrument. For example, let us consider the following item:

Which of the following jobs *requires* the most education?

a. college professor
b. accountant
c. garbage collector
d. teacher

Suppose that an analysis of 100 responses to the item revealed the following pattern:

a. 20 individuals
b. 40 individuals
c. 0 individuals
d. 40 individuals

Obviously, response "c" is not effective in helping to distinguish among student reactions; therefore "c" should be replaced by a more realistic response. Discussing the item with students might generate a more realistic alternative (for example, "stock broker"). The instrument should then be readministered, and the number of responses should once again be tabulated for each item. Let us imagine that the individual responses for this item would then be:

a. 30 individuals
b. 20 individuals
c. 15 individuals
d. 25 individuals

All of the responses now appear to be more realistic, and they all help distinguish individuals who responded correctly from those who didn't. Therefore, the substitution would make the instrument more reliable.

3. **Retaining items that differentiate higher and lower scoring individuals.** Another way to increase the reliability of an instrument is to include only those items that the higher scoring individuals answer correctly and the lower scoring individuals answer incorrectly. To determine these items, arrange the students by scores with the highest score first and the lowest score last. Next, divide these students into four equal groupings so that you can place the highest 25 percent into one group and the lowest 25 percent into another group. Then examine the instrument item by item as shown in Figure 16.

Figure 16

A COMPARISON OF HOW THE TOP AND BOTTOM GROUPS RESPONDED TO ITEMS IN A CAREER EDUCATION INSTRUMENT

Item	Top 25			Bottom 25		
	No. of right responses	No. of wrong responses	No. not responding	No. of right responses	No. of wrong responses	No. not responding
1	20	5	0	4	19	2
2	25	0	0	3	21	1
3	15	6	4	7	16	2
4	23	0	2	5	18	2
5	22	3	0	6	16	3
6	5	20	0	20	4	1
7	21	4	0	4	20	1
8	22	3	0	11	12	2
9	3	20	2	21	3	1
10	24	1	0	5	19	1

In this example, a total of 100 individuals responded to the instrument, so the top 25 and the bottom 25 are included in the analysis. For each item, the number of correct responses, the number of incorrect responses, and the number not responding are reported for each group (top 25 percent and bottom 25 percent). In this example, items 6 and 9 should be changed. Those individuals who scored high on this exam tended to answer the items incorrectly, whereas those who scored low tended to answer them correctly. Once again, discussing the items with the class will probably clarify their responses and present alternative items.

4. Lengthening the test. Generally, the length of an instrument is closely related to its reliability. So in developing an instrument you should be aware of the two dimensions of length: number of items and necessary completion time. You should probably write twice as many items as you plan to use, so that after you have tested the instrument you can substitute the better items for unsatisfactory ones. The time required to complete the instrument should not exceed the amount of time for which the individuals would normally be expected to concentrate and remain still. For example, an instrument for secondary students should have a completion time of no longer than a regular class period (approximately forty-five to fifty minutes). An instrument for elementary students would normally be shorter. Also, you should allow enough time for most students to complete the test; naturally some will finish before others.

5. Providing clear directions. If they are to provide reliable responses, individuals taking a career education test must know what is expected of them, what the instrument is designed to do, how long it will take to complete, whether they should guess when responding to a question, and what is to be done with the results. Therefore, the directions must be thorough and detailed.

Compare the two sets of directions in Figure 17. Which of the two directions would you rather receive if you were taking the test? The second set is thorough and explicit and should inform the students what is expected of them. For the sake of clarity, it may even be advisable to read the directions aloud before the testing period begins.

One consideration in drafting directions is the fact that some school districts and state departments of education are requiring explicit procedures for ensuring privacy and the protection of testing subjects. You should become familiar with these regulations before you develop your career education instrument. The checklist provided in Figure 18 is designed to guide your development of effective directions.

6. Eliminating ambiguous items. Ambiguous items are often misunderstood, therefore, they can cause an instrument to be unreliable. If an instrument is to be reliable, it is essential that the items convey precisely what you mean to communicate. Several factors can create ambiguity in test items. Four are discussed below:

- *Use of homonyms* — A number of words resemble one another in sound and spelling but differ in meaning (such as "sublimate," to divert the energy of an impulse to a higher purpose, and "sublimate," to convert a solid into a vapor which on cooling condenses to a solid again). Other words sound alike but are spelled differently and mean different things (such as "see" and "sea," or "bare" and "bear"). All such words can mislead students taking a test and thus should be used with care.

- *Dialect differences* — The word "bright," for example, can be used to mean "intelligent," but blacks also interpret this word as referring to someone of light skin pigmentation. Similarly blacks can interpret the word "bad" as "bad" or as "good" depending upon its usage. Additionally, some words do not translate into other languages, for example the term "coping" does not translate into a Spanish equivalent.

Figure 17

TWO SETS OF DIRECTIONS FOR A CAREER EDUCATION INSTRUMENT

Example A

DIRECTIONS: Please read each item carefully and make your response as accurately as possible.

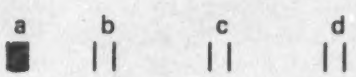
Example B

DIRECTIONS: This test is designed to determine your knowledge of the world of work. The results of this test will be used to determine the effectiveness of the social studies instruction in which you participated throughout the school year. The results will also represent 25 percent of your final grade in social studies. Please print your name in the space at the upper right hand corner of the answer sheet. This test contains thirty multiple-choice questions. Read each question carefully and choose the *one* best answer to each question. *On your answer sheet* darken the space underneath the letter that matches the answer you choose. Read the sample question below to see how the answer is to be marked on your answer sheet.

SAMPLE: Which of the following tools would a carpenter be most likely to use?

a. square
b. trowel
c. wire cutters
d. volt meter

(a) is the correct answer. On your answer sheet you should darken the space under the letter a:



You will receive two points for each correct answer. Do not hesitate to guess if you do not know the correct answer. You will have thirty minutes to complete this test. When you have finished, return both the test and answer sheet to your teacher.

Figure 18

A CHECKLIST FOR IMPROVING DIRECTIONS FOR CAREER EDUCATION INSTRUMENTS

Yes	No	
☐	☐	1. Are individuals told how to identify themselves on the tests or answer sheet?
☐	☐	2. Do the directions include sample items that illustrate how to mark a response?
☐	☐	3. Have individuals similar to those taking the tests been given an opportunity to react to the directions?
☐	☐	4. Do the directions explain the purpose of the test?
☐	☐	5. Do the directions specify a time limit for completing the test?
☐	☐	6. Do the directions state where the individual is to respond (e.g., on the test or on an answer sheet)?
☐	☐	7. Do the directions indicate whether the individual should or should not guess?
☐	☐	8. Do the directions explain how the results will be utilized and reported?
☐	☐	9. If specialized (optical scan) response sheets are required, are individuals told to avoid stray marks?
☐	☐	10. Are individuals told how to change a response if they wish to do so?
☐	☐	11. Are individuals told how the test will be scored?
☐	☐	12. Are the sentences short and clear?
☐	☐	13. Are the directions readily visible to the individual taking the test?
☐	☐	14. Are the individuals told what to do when they have completed the test?
☐	☐	15. Have you checked with local agencies regarding the policy for data collection from people?
☐	☐	16. Have you checked with state agencies regarding policies for data collection from people?

- *Compound predicates or "double barreled" sentences* — An item could elicit two responses when the instrument allows only one response per item, as in the following example:

> I like to keep busy and complete tasks on time.

The respondent may indeed like to complete a task on time but not keep busy. How is he or she to respond? This item should be divided into two separate items as follows:

> 1. I like to keep busy.
> 2. I like to complete tasks on time.

- *Inappropriate use of words* — Items sometimes contain poor sentence structure or incorrect grammar and, at other times, items are stated in a wordy manner. To avoid these pitfalls, once you have developed an item, read it aloud, listen or have someone else listen to the item. You might also want to have someone read your items to see if you have overlooked any errors. It is not unusual for people to overlook their own errors because they know what they meant to say. However, someone who has to rely on the printed word when trying to understand what you mean is more likely to spot possible problems.

The six suggestions presented in this unit should assist you in improving the reliability of your career education instrument. Once you have refined your instrument by using these techniques, you may want to have the reliability checked by the more sophisticated procedures employed by an expert. If so, you should be able to obtain assistance from your district or state department of education; from colleges and universities, with educational laboratories and centers; or from private educational consultants and agencies. While we have noted ways to improve reliability, we should not leave this topic without a note of caution. Changes made to improve the reliability of your career education instrument affect the validity of the instrument if the item no longer represents the area that is to be assessed. (Validity will be discussed in the next chapter.) Therefore, changes made to improve reliability should be made with care.

DETERMINING VALIDITY

The adage that states "you cannot judge a book by its cover" could well be extended to read "you cannot judge a test by its title." For example, an instrument entitled "Career Decision Making" may not actually constitute a test on this activity. It may be simply a collection of quickly assembled items loosely related to the topic and rushed into use. This section of the handbook is designed to help you understand what is meant by the term validity, how you can check your instrument's validity, and how you can improve the validity and therefore the quality of your own career education instruments.

What Is Meant by the Term Validity?

Validity refers to the accuracy with which a test measures what it is supposed to measure. The literature on testing generally describes four major types of validity: content, concurrent, predictive, and construct. You need to determine which of these types are most appropriate for testing your instrument.

Content validity refers to the representativeness of the test items and their relationship to the topic being measured. For example, a test on decision making would not be able to include *all* the information needed by a student, but it should contain an adequate sampling of information from the content area. Furthermore, the test should exclude information on tangential areas, such as the world of work, economic understanding, or self-awareness.

Concurrent validity determines how much agreement exists between a score on a particular instrument and a score obtained from other acceptable instruments, bodies of knowledge, or skills in the same content area. An example of concurrent validity would be the similarity between the students' scores on an instrument on work habits and their actual habits observed on the job.

Predictive validity is the ability of an instrument to predict future performance based on predetermined criteria. For example, scores on the Strong-Campbell Vocational Interest Test have been shown to predict later career choices. The major difference between predictive and concurrent validity is the time element: concurrent validity is demonstrated in the present, while predictive validity is demonstrated in the future.

Construct validity "proves" certain hypothetical relations. The construct validity of a career interest instrument, for example, determines what extrinsic factors (such as general knowledge of careers, attitudes toward work, and views toward the locus of authority for personal development) account for the variations in the test scores. The hypothetical nature of construct validity differentiates it from other types of validity and places it among the concerns of researchers and professional test developers; therefore, construct validity will not be discussed in this handbook.

This section has discussed three different ways to determine the validity of your career education instrument. The next section presents specific suggestions for improving the validity of your instrument.

How Can I Check a Career Education Instrument for Validity?

This section will discuss how you can determine three of the preceding types of validity: *content, concurrent,* and *predictive.*

Content validity can be evaluated by having experts review and rate the test items for relevance. A panel review by three to nine authorities is desirable. For example, if you were working on an instrument gauging economic understanding, economists should review the test to determine if it does in fact measure economic understanding. In this review process, the experts should be asked to respond to such questions as:

- Are all the items related to economic understanding?
- Do the items provide an adequate assessment of economic awareness?
- What items should be added to make the instrument more representative?
- Does the title of the instrument accurately represent the material included?

A sample form that could be used in this process is presented in Figure 19.

Concurrent validity can be determined by comparing the results obtained from your instrument with an independent assessment of a different, previously validated instrument. These two measures should be obtained at the same point in time. If there is agreement between these two separate instruments, your instrument is said to have concurrent validity with the previously validated one.

An easy method for establishing concurrent validity is to arrange the students' names in sequential order from the highest score to the lowest on each of the two separate instruments. Once the students have been ranked in order, you should proceed to compare the results of your instrument with the results of the previously validated instrument. In making this comparison you would use the same procedure suggested for determining the equivalence of two forms of the same test that was discussed in the reliability section (see p. 24). In determining concurrent validity the names of each student on the previously validated test would be placed in rank order (from highest to lowest) and the appropriate rank for each student would be placed in a column to the right of the appropriate name column. The rank order for each student on your test should then be placed in a second column to the right of the appropriate name (see Figure 14, p. 25). You can then proceed to compute the concurrent validity following the same formula and procedures discussed for the equivalence reliability. The closer this number is to 1.00 the more valid your test is said to be.

Predictive validity is determined by comparing the present scores on an instrument with future scores from a different instrument. For example, let us assume that you have developed an instrument on occupational awareness and have administered it to twelfth grade students. In general, those who are more knowledgeable about occupations tend to make better career decisions, so you want to determine if the students who scored higher on your instrument will be satisfied with their educational and career choices. Therefore, an independent test of job satisfaction administered to these same students at a later time could be used to determine if the occupational awareness instrument did, in fact, forecast job satisfaction.

In general, you can determine predictive validity in much the same way that you determine both concurrent validity and equivalence reliability. First, you order the names, by score, from

Figure 19

A SUGGESTED FORM FOR A PANEL REVIEW OF AN ECONOMIC UNDERSTANDING INSTRUMENT

DIRECTIONS: From your experience, please judge whether the items included on the Economic Understanding Instrument are necessary for high school students to know in order to make an effective transition from school to work. In making this judgment, please check your opinion (yes, no, or undecided) on the scale to the right of each numbered item. Any comments regarding this item should be written in the space to the far right.

Item	Yes	No	Undecided	Comments
1	☐	☐	☐	
2	☐	☐	☐	
3	☐	☐	☐	
4	☐	☐	☐	
5	☐	☐	☐	
6	☐	☐	☐	
7	☐	☐	☐	
8	☐	☐	☐	
9	☐	☐	☐	
10	☐	☐	☐	

In the following space (and, if necessary, on the back of this page) please comment on the entire test.

1. Is the content of this instrument representative of the knowledge required of high school students to make more effective transitions from school to work? If not, what else should be added?

2. Are all the items related to economic understanding? If not, which ones should be eliminated?

3. Does the title of this instrument represent the material included in the individual items? If not, what should the instrument be called?

highest to lowest on the first (occupational awareness) instrument and place the number of the appropriate rank order in a column to the immediate right of each name. This column should be in correct numerical order. You then proceed to rank the students' names in order (from highest to lowest) according to their scores on the second (job satisfaction) test. The rank of each student on the job satisfaction test is then placed in a second column to the right of the appropriate name (see Figure 14, p. 25). Again, the formula and procedure for determining the predictive validity of a test is the same as that suggested for equivalence reliability (see p. 24). Remember, if you want to determine predictive validity you must plan ahead and retain test scores on one test so that you can compare them with scores from another test at some future point in time.

How Can I Improve the Validity of a Career Education Instrument?

This section provides six practical suggestions for improving the validity of your career education instrument:

1. Prior to revising the instrument, review the information on the goals and objectives of the career education topic to be assessed.

2. Use a planning matrix to develop the instrument.

3. Research occupations so that you can develop items based on the different jobs that various individuals perform.

4. Develop twice as many items as you intend to use in the instrument.

5. Lengthen the instrument.

6. Have experts review the instrument for relevance to the topic being assessed.

1. Reviewing information on the goals and objectives. Please refer to Unit 1, "Planning and Designing," pp. 9-10.

2. Using a matrix. Using a planning matrix (see Figure 7, p. 17) to develop a career education instrument is likely to increase the validity of the instrument for several reasons. First, the matrix facilitates the development of a clear plan for the instrument. Second, the matrix enables you to check the relevance of test items to the overall plan and thus keep the test "on target." Third, the matrix facilitates the selection of a representative number of items to achieve the purpose of the instrument. And fourth, the matrix enables you to regulate the kinds of responses called for by the test items; this advantage is important because items that require an individual to respond in ways other than memorization tend to constitute more valid instruments. Using a planning matrix, then, can be an important aid to improving validity in an instrument.

3. Researching occupations. Educators seldom have sufficient information about occupations to develop valid career education instruments. Therefore, it is frequently necessary to research the kinds of work required in various occupations. One of the most obvious ways to do this is to talk to someone in that occupation. Another way is to read about various occupations. Fortunately, several readily accessible reference works can help supply this background information:

- *The Dictionary of Occupational Titles* (1965)

- *The Occupational Outlook Handbook* (revised every few years)

- *Directory of Task Inventories* Volume I (1974), Volume II (1975), and Volume III (1976)

The more precisely your items present the real world of work, the more valid your instrument is likely to be.

4. Devising sufficient items. A good rule of thumb to follow when developing an instrument is to devise twice as many items as you intend to use. If you can develop a "pool" of items on the topic you wish to assess, you can then draw upon these items to develop an instrument that is more representative of the topic. The fewer the items that you have, the less likely that instrument will assess the topic thoroughly. You can then categorize these test items by using a planning matrix (such as that illustrated in Figure 7 on page 17) in order to determine if appropriate emphasis has been placed on each of the learning tasks and domains included in the instrument.

5. Lengthening the instrument. Though it is overlooked, lengthening the instrument is a good way to improve its validity. Including more items allows you to sample more of the learning tasks that need to be included. Consider the example of an instrument on work habits that consists of five learning tasks (the state of Florida, 1973): time utilization, responsibility, safety and neatness, use of instruction, and attitude toward the job. In an instrument consisting of twenty-five items with equal emphasis on each learning area, no more than five items could be included for each learning task. With this limited number of items, the results may not reflect an individual's work habits. Doubling the number of items for each learning task increases the likelihood that the results would more accurately reflect the individual's work habits. Lengthening the instrument, then, can increase its validity. However, also keep in mind the length of time it takes to administer a test—completion time often limits the number of items that can be included on an instrument.

6. Reviewing by experts. Once you have developed the initial draft of a career education instrument, you should have knowledgeable people (e.g., teachers, subject matter specialists, business persons, etc.) review the proposed test. This review, sometimes referred to as assessing the "face" validity, could follow the same format as that suggested for content validity (see Figure 19, page 35). A second review technique is having the reviewers classify the test items according to the learning tasks or classes of the appropriate educational domain. Educators may classify more effectively the items according to the class of the educational domain being considered, whereas business persons and other noneducators may classify more effectively the items according to the learning tasks. In classifying test items you could use a modification of the planning matrix (page 17) and ask each person to write the number of the item under the appropriate heading. An example of how this planning matrix could be modified to record the number of test items by class of educational domain is presented in Figure 20. An example for classifying items by content area is shown in Figure 21.

Figure 20

A SAMPLE FORM FOR RECORDING TEST ITEMS BY CLASSES OF THE COGNITIVE EDUCATIONAL DOMAIN

	Classes of the Cognitive Domain					
	Knowledge	Comprehension	Application	Analysis	Synthesis	Evaluation
Major Content Areas						
Total Number of Items						
Percent of Total Items						

Figure 21

A SAMPLE FORM FOR RECORDING TEST ITEMS BY CONTENT AREA

	Content Area					
	General Procedures for Applying for a Job	Job Interviews: General Interviewing Procedures	Job Interviews: Manners and Appearance	Job Interviews: Communication Skills	Application Forms and Resumes	Pre-Employment Testing
Major Content Areas						
Total Number of Items						
Percent of Total Items						

ELIMINATING STEREOTYPES

Career educators in particular should be concerned with breaking down stereotypes so that all individuals will be encouraged to pursue meaningful careers. As with any curriculum, career education materials—not only instruments but also accompanying manuals, resources, and reporting formats—should be free from stereotyping of any type. Most publishers and agencies in the field of education have guidelines for authors to follow in preparing materials that are free from stereotyping. This section of the handbook, based upon these widely accepted guidelines, is designed to help you reduce sex and ethnic stereotyping in career education instruments.

Are My Career Education Instruments Free of Stereotypes?

This section presents five practical suggestions for reducing stereotypes in career education instruments.

1. Watch your pronouns. The pronoun "he" has often been used in a generic manner; today, however, this usage is considered abrasive and tends toward the continuation of sexual stereotypes. One way of eliminating the stereotype is to substitute the terms "he/she," "her/him," "himself or herself," "hers or his," etc. In addition, you should vary the sex of the pronoun given first (i.e., female or male). This remedy should not be overused, however, as it can become obtrusive. A less obtrusive way of eliminating the stereotype is to change singular nouns to plural nouns and then use the accompanying plural pronouns "they," "theirs," etc. There are also other ways to avoid the generic "he": the noun can be repeated, and sentence structure can be changed to avoid the dilemma.

2. Use neutral terms in describing various occupations or careers. Many words referring to workers and professionals reflect a male-dominated society. These words include "businessman," "chairman," "mailman," "fireman." Listed below are neutral alternatives for several job designations:

Previous Usage	Neutral Term
Businessman	Businessperson
Chairman	Chairperson
City hostess	Goodwill ambassador
Fireman	Fire fighter
Maid	Houseworker
Mailman	Mail carrier
Manpower	Labor force
Policeman	Police officer
Salad girl	Salad maker
Stewardess	Flight attendant

The U.S. Department of Labor has published a document entitled *Job Title Revisions to Eliminate Sex- and Age-Referent Language from the Dictionary of Occupational Titles* (1975). If you are unable to think of a neutral term to describe an occupation, this publication may be helpful.

3. Include in items on the instrument names that are different from Anglo in origin. The tradition of using only Anglo names in educational materials has been cited as one reason that minorities may not have seen the full range of career opportunities that others have seen. Therefore, it is important that career education instruments use names that are of obvious ethnic origin, such as:

Coretta	Juan
Yumiko	Nariyo
Mohammed	Rao
Inez	Gilda
Zuynk	Carita
Youngja	Sugil
Shoshana	Samuel

Undoubtedly, you will be able to list many others simply by drawing upon the names of celebrities, colleagues, and friends.

4. Include items that reflect a variety of life styles, personality types, socio-economic levels, cultures, and handicaps. Educational materials have long mirrored the "middle class" value system of Americans. These materials invariably depicted a nuclear family with the father as breadwinner and the mother as homemaker. Seldom were extended families or "broken homes" depicted. Similarly, little attention was given to different cultures or socioeconomic levels. Furthermore, handicapped persons never appeared in educational materials. Many individuals using these materials naturally found it difficult to identify with the characters and consequently their educational performance was hampered. Therefore, career education instruments, whenever possible, should illustrate a variety of life styles, personality types, socioeconomic levels, cultures, and handicapping conditions. In presenting more varied life styles, however, you must be careful to avoid perpetuating sex-linked or ethnic-linked stereotypes (e.g., outgoing males, shy females, black broken home).

5. Include some items that depict males in traditionally female roles and occupations and females in traditionally male roles and occupations. The news is full of accounts of both men and women excelling in activities that were once thought to be the exclusive domain of the "opposite" sex. Consequently, illustrations of men and women in wider ranges of roles and occupations not only encourage equal rights and opportunities, but also reflect today's world. Career education materials, because of their special impact on future conditions within professions, should depict a broad range of opportunities for both men and women. The materials should also include blacks in executive or supervisory roles and role "reversals" (e.g., white working for black, man working for woman).

DEVISING THE FORMAT

First impressions are lasting. Getting beyond a negative first impression to a more thorough and fair appraisal generally takes quite some time. If the first impressions generated by a career education instrument are negative, the persons taking the test may not begin to perform at their highest levels. Therefore, in order to develop career education instruments that provide accurate and reliable results, it is advisable to devise a format that makes a good first impression by means of a clear, neat, logical, and efficient layout.

How Can I Improve the Format of My Career Education Instrument?

This section is designed to provide a list of four suggestions for designing or improving the format of your career education instrument.

1. Limit the number of different types of items on your career education instrument. Each time that you change the type of item included on your career education instrument (e.g., true-false, multiple choice, short answer, matching, essay, rating scale, etc.), you can disorient the individuals taking the test. Therefore, it is desirable to limit the number of different types of items on tests and make clear transition from one to another when necessary. New directions should be included whenever the format changes. A good rule of thumb is to use only two or three different types per instrument. Similarly, it is helpful to group all items of the same type together. For example, on an instrument with three types of items, keep all the multiple choice items together, all the matching items together, and all the true-false items together. This grouping will enable the examinees to lose less time reading directions and reorienting themselves to the format, and thus allow more time for responding to the items.

2. Use large type and clear illustrations when preparing the instrument. When taking a career education test, individuals should have to concern themselves only with reading the question and making an appropriate response. It is counterproductive and in fact, unfair to require these individuals to puzzle over blurred words or complicated illustrations in addition to providing a response to the item. Therefore, you should always use large, clean type or print when preparing a career education instrument. This suggestion is particularly important as you will subsequently have to reproduce the instrument. The higher the quality of the master copy, the higher will be the quality of subsequent reproductions. Make responses as short as possible, however, in doing so do not sacrifice clarity.

Illustrations also need to be clear. Unnecessary information and detail should be eliminated so that individuals can quickly focus their attention and respond to the item. In addition, an illustration should appear on the same page as the items that refer to it so that individuals need not lose time flipping pages back and forth.

3. Keep all possible responses to an item on the same page. An important guideline to remember when preparing a career education instrument is: simplify the response format. One way to do

this is to include all possible responses to an item on the same page as the item. Imagine the difficulty the respondents would have if they had to flip back and forth between pages in order to consider multiple choice items C and D or to complete matching items in an exercise.

4. Prepare an attractive page layout. The logical sequence of exercises and the attractiveness of entire pages are focal points for preparing an attractive layout. If blanks are required for completing items, they should be of uniform length. If the blanks are placed to the left of the items, they will not only add regularity and neatness to a prominent visual portion of the page but will also make scoring easier. Similarly, placing all possible responses directly beneath an item will improve the attractiveness of the layout. It is necessary, though, to consider the responses for all of the items on a single page (both multiple choice and matching, for example) and to select the pattern that allows for consistency as well as neatness of design.

The appearance of the entire page is very important. Items should be spaced so that the page appears uncluttered. Ideally, margins should be uniform on both sides of the page. The balance between printed material and unprinted space should be even and symmetrical. Even after following these guidelines, however, you should have a group of teachers review the layout and offer their suggestions for improvement.

GAUGING READABILITY

An often overlooked criterion for judging your career education instrument is whether respondents are able to understand what you have written. This criterion is particularly important when the respondents are of vastly different backgrounds and abilities. The instrument need not be directed at the lowest levels of achievement in order to test all the respondents.

How Can I Improve the Readability of a Career Education Instrument?

This section will present three suggestions for improving the readability of a career education instrument: (1) develop the instrument so that it progresses from a few relatively simple items at the beginning to more complex items later on, (2) conduct readability tests, and (3) pilot test the instrument with individuals similar to those for whom you are preparing the instrument.

1. **Begin with relatively easy items.** Accepted principles of educational psychology dictate that teachers should progress from the familiar to the unfamiliar and from the easy to the difficult. These principles also apply to the development of career education instruments. Therefore, you should gauge the level of difficulty for all of the items, either by classifying them yourself (e.g., as easy, moderate, or hard) or by conducting a pilot test (as described in suggestion 3). After gauging the difficulty of the items, you could then organize them by beginning with easier items and progressing to the more difficult ones.

2. **Conduct readability tests.** The reading level of a career education instrument can be obtained by using the readability formulas found in numerous education texts. Dale and Seels (1971) have developed a comprehensive bibliography on readability measures. Other sources include Flesch (1951), Bormouth (1969), and Chall and Dale (1948). However, it should be noted that formulas alone do not furnish exact answers. You should still conduct a pilot test of an instrument prior to its use.

3. **Pilot test the instrument.** The reading level of a career education instrument should not be condescending or childish; nor so difficult that it discourages examinees and thus undermines the results obtained from the instrument. One way to check the readability is to conduct a pilot test of the instrument by assembling a representative group of individuals and asking them to identify the major problems they find in understanding the instrument. This method is less sophisticated than the complex methods noted in suggestion 2. This group should represent as wide a range of academic, socioeconomic, and cultural levels as possible and should include both males and females. After each of the reviewers has marked all the difficult words and items, ask the group to discuss the instrument on an item-by-item basis and indicate any problems that they find. In working with younger groups, you should lead the students through the process one section at a time. If you ask the individuals to indicate their reactions independently of one another, you minimize peer group pressures; thus you improve your chances of noting different points of view and of providing a more realistic review of the instrument.

EXAMINING YOUR CAREER EDUCATION INSTRUMENT

The preceding units of this handbook have presented six topics related to improving career education instruments: planning and designing, checking reliability, determining validity, eliminating stereotypes, devising the format, and gauging readability. These units have provided you with a basic awareness of some problems encountered in developing instruments as well as some techniques for minimizing these problems. With this basic awareness, you are now ready to examine the instrument that you have developed. Your goal is to establish confidence in your instrument, two areas of major concern in such a task are those of validity and reliability. As a start, let's take a look at the content validity and internal consistency reliability of your instrument.

Determining Content Validity

First, let us determine the instrument's content validity. In making this determination it will be necessary for you to develop and utilize the format and procedure for conducting a panel review. A sample format was suggested on page 35. Review this format and procedure and then modify the format to meet the needs for your instrument. You will probably have to change a few words in the directions and questions asked and adjust the form for the number of items on your instrument. Once you have made these modifications, select a panel of individuals that are considered knowledgeable of the area to be assessed. It would be desirable for you to include persons who are knowledgeable from education as well as personnel from business, industry, and labor.

In conducting the panel review, it is generally desirable to have each member of the panel, review the instrument independently. Once they have completed their reviews and completed the panel review forms, ask panel members to share their reactions with each other. This discussion should provide the information you need to determine how well your instrument measures what you had intended for it to measure. If most of the panel members are basically pleased with the instrument you are ready to proceed. However, if the panel is divided in its opinion or if most of the panel members are unhappy with the instrument, you will need to do some additional work on the instrument before moving on to the next step. If additional work is required you should conduct a second panel review of the revised instrument to make sure necessary improvements have been made before proceeding to next step: determining internal consistency reliability.

Determining Internal Consistency Reliability

If the panel has decided that your career education instrument measures what you intended for it to measure, you are ready to determine its internal consistency reliability. In making this determination it would be advisable to administer your instrument to a minimum of thirty students. (Ideally, you should check the reliability of an instrument with ten times as many students as you have items on a test. For example, if your instrument has twenty-five items, it would be desirable to administer it to 250 students.) You also need to make sure that these students are representative of the intended population for the instrument. Every effort should be made to see that these students represent all cultural, economic, educational, and ethnic backgrounds. The sample should also

include the disadvantaged, the handicapped, and equal number of representatives of both sexes. After a sufficient number and variety of students have been selected and the test has been administered you can proceed to determine its reliability.

First, it will be necessary to divide the test into two parts. An acceptable way of doing this is on the basis of odd- and even-numbered items. Score each of these two sections, determining a final score for each. Once these two scores are computed for each student you are ready to begin computing the internal consistency reliability.

Second, arrange the tests in rank order from the highest score to the lowest score for each student on the even-numbered items. Now, prepare a list of these students' names in rank order from the highest to lowest score and number them accordingly.

Fourth, merge the information on these two lists into one form as shown in Figure 12 in the reliability section. A sample form for your instrument is shown in Figure 22.

Fifth, compute the internal consistency reliability using the formula:

$$T = \frac{S}{\frac{1}{2} N (N-1)}$$

Substitute the appropriate figures for your instrument and complete the computation. A sample problem of this type is shown on pages 22-23. The value you compute for T should be between 0.00 and +1.00. If this is not the case, you have made an error and you should recheck your computations. If your answer is within the range mentioned above, check with the guidelines presented on page 19 for interpreting reliability coefficients. How acceptable is your reliability coefficient? If it is not at least .80 (see Figure 8) you should give additional thought to revising your instrument. Specific suggestions for increasing the reliability of tests are presented on pages 26-32.

Figure 22

A SAMPLE FORM FOR RECORDING SCORES WHEN DETERMINING INTERNAL CONSISTENCY RELIABILITY

Name	Rank on Odd-Numbered Items	Rank on Even-Numbered Items
	1	
	2	
	3	
	4	
	5	
	6	
	7	
	8	
	9	
	10	
	11	
	12	
	13	

SOME CONCLUDING REMARKS

Developing sound evaluation instruments is challenging. Educational measurement is a relatively new discipline, and although it is gaining sophistication and acceptance, it will always be impossible to gather all of the information available in any attempt to measure human activity. The task of developing an instrument for the field of career education is further complicated by the fact that career education is also a new discipline, and it is still developing the precise definitions and concepts that characterize the more established scientific and technical fields.

A number of books and articles have been prepared to assist in the measurement of human activity. Certainly these materials, although not developed specifically for career education, have relevance to career education and a number of suggestions can be obtained through their use. We have cited a number of sources throughout this handbook, and while it would be difficult to list all of the materials that would be useful to you, we do think two books of general use merit your attention:

Teacher-Made Tests (Green, 1975)

Measuring Educational Achievement (Ebel, 1965)

Finally, this handbook was designed to help allay some of your fears and concerns about developing evaluation measures. Because of the diverse needs of each career education program, the handbook was designed to identify major basic issues and offer suggestions for you to consider when developing measures. Specifically, it addressed the topics of improving the planning and designing, checking reliability, determining validity, devising the format, examining the readability and eliminating stereotypes of career education instruments. Another handbook in this career education measurement series, *Career Education Measures: A Compendium of Evaluation Instruments,* also provides examples of other career education instruments that have been developed by other individuals. Thus, the materials contained in this handbook and the suggested references together should provide you with a foundation for designing improved career education measures. By using these sources and others that you may be familiar with, you should now be ready to develop and improve career education measures to meet your needs. We think you will find this to be a challenging and rewarding activity.

REFERENCES

Basic Learner Outcomes for Career Education. Austin, Texas: Texas Education Agency, 1977.

Bloom, Benjamin S. and others. *Taxonomy of Educational Objectives: The Classification of Educational Goals. Handbook I: Cognitive Domain.* New York: David McKay Company, Inc., 1956.

Bormouth, John R. "Readability: A New Approach," *Reading Research Quarterly* 1 (Spring 1969).

Career Education: Career Development Goals and Performance Indicators. Lansing, Michigan: Michigan State Department of Education, 1974.

Chall, Jeanne and Dale, Edgar. "A Formula for Predicting Readability," *Educational Research Bulletin* 27 (January 21, 1948).

Cronbach, Lee J. *Essentials of Psychological Testing.* 3rd ed. New York: Harper and Row, Publishers, 1970.

Dale, Edgar and Seels, Barbara. *Readability and Reading. An Annotated Bibliography.* ED 075 789. Newark, Delaware: International Reading Association, 1971.

Developmental Program Goals for the Comprehensive Career Education Model, Preliminary Edition. Columbus, Ohio: The Center for Vocational and Technical Education, The Ohio State Univerwity, August 1972.

Dictionary of Occupational Titles. Vol. I. Definitions of Titles. 3rd ed. Washington, D.C.: Manpower Administration, Department of Labor, 1965.

Directory of Task Inventories, Volume I. Columbus, Ohio: The Center for Vocational Education, The Ohio State University, 1974.

Directory of Task Inventories, Volume II. Columbus, Ohio: The Center for Vocational Education, The Ohio State University, 1975.

Directory of Task Inventories, Volume III. Columbus, Ohio: The Center for Vocational Education, The Ohio State University, 1976.

Ebel, Robert L. *Measuring Educational Achievement.* Englewood Cliffs, New Jersey: Prentice Hall, Inc., 1965.

"Eighth Annual Gallup Poll of the Public's Attitude Toward the Public Schools." *Phi Delta Kappan* 58 (October 1976).

Evaluation and Educational Decision Making: A Functional Guide to Evaluating Career Education. ED 117 185. Washington, D.C.: Development Associates, Inc., 1975.

Flesch, Rudolph. *How to Test Readability.* New York: Harper and Brothers, 1951.

The Florida Catalog of Occupational Objectives: Employability Skills. Tallahassee, Florida: Florida Department of Education, 1973.

Galladay, Mary A. *The Condition of Education: 1976 Edition.* Washington, D.C.: National Center for Education Statistics, U.S. Department of Health, Education and Welfare, 1976.

Green, John A. *Teacher Made Tests, Second Edition.* New York: Harper and Row, Publishers, 1975.

Hoyt, Kenneth B. *Perspectives on the Problem of Evaluation in Career Education.* Monograph on Career Education. Washington, D.C.: Department of Health, Education and Welfare, 1975.

Job Title Revision to Eliminate Sex- and Age-Referent Language from the Dictionary of Occupational Titles, Third Edition. Washington, D.C.: Manpower Administration, U.S. Department of Labor, 1975.

Katz, Martin. *Selecting an Achievement Test: Principles and Procedures.* Princeton, New Jersey: Educational Testing Service, 1973.

Kibler, Robert J.; Larry L.; Parker; and Miles, David T. *Behavioral Objectives and Instruction.* Boston, Massachusetts: Allyn and Bacon, Inc., 1970.

Krathwohl, David R.; Bloom, Benjamin S.; and Masia, Bertram B. *Taxonomy of Educational Objectives: The Clarification of Educational Goals. Handbook II: Affective Domain.* New York: David McKay Company, Inc., 1956.

Occupational Outlook Handbook, 1976-77 Edition. Washington, D.C.: U.S. Department of Labor, Bureau of Labor Statistics.

Siegel, Sidney. *Nonparametric Statistics for the Behavioral Sciences.* New York: McGraw Hill Book Company, 1956.